iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 11 **Issue:** V **Month of publication:** May 2023

**DOI:** https://doi.org/10.22214/ijraset.2023.51955

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 11 Issue V May 2023- Available at www.ijraset.com*

# Machine-Learning Approach for Predicting Harmful Diseases using Big Data and IoT: A Review

Ragini Pandey[1], Dr. Kamal Kumar Shrivastava[2]
[1]*Student,* [2]*Associate Professor, Shri Krishna University, Chhatarpur(M.P) India*

***Abstract:*** ***The magnitude of the enormous quantity of data in the healthcare sector is currently on the order of petabytes, and it is expanding exponentially. The massive increase of data has resulted in a number of issues with computing, storage, and transfer. With the aid of a conventional relational database system, this type of data may be evaluated and processed. Moreover, structured data can only be processed by standard database systems. Big data, on the other hand, stores unstructured data. ICT would be able to assist society in a better storage with invention of new and effective mechanisms for the storage and accessing of the information. ICT implementation in the healthcare sector is referred to as eHealth. The processing of data, subsequent analysis, and enhancement of the decision-making processes would all benefit from application in the health care sector in order to gather better treatment options for diseases' symptoms.***

## I. INTRODUCTION

An important component of many applications is data analytics. Due to the increased amount of heterogeneous & unstructured data available globally, data analytics have become an essential tool for scientists. Because enormous amounts of data are frequently aggregated across several businesses in the medical sector, scalable data analytics approaches are required. These data sources serve as a resource for gaining insights about how to improve healthcare service and reduce waste. Analysis and application of such data in a real-world healthcare setting are difficult due to the volume and complexity of the data. Massive volumes of data, which are generated, acquired, stored, & analyzed every second. The use of the Internet, which enables data transmission between different technological devices and people, has led to a revolution with in generation of huge volumes of data. Mobiles, social media, imaging technology used to make medical diagnoses, etc. are all referenced in this context. The amount of data that is readily available keeps expanding, and it does so in various formats. On the other hand, the cost of data storage is continuing to decrease, making it more affordable. Although it is becoming more affordable and accessible to build data storage, the growing volume of data in various formats and from many sources poses new challenges for data processing, including its analysis and the incorporation of Big Data in business decision-making processes. New technologies are being developed to solve these issues in order to process and store Big Data. A fresh strategy is required to address these problems, such as creating an elastic and scalable infrastructure. The goal of this study is to investigate the field of the big data problem, in particular, to create an overview of the free biomedical big data repositories that are currently available and to identify appropriate technologies and methods, along with their advantages and disadvantages, to be used with the selected data. A test scenario is developed and deployed on this data because the data, technology, and methodologies have been selected. The achieved results then are discussed, and some suggestions are made.

Even while e-health is still a relatively new idea globally, over the past few decades it has become more popular in many nations. Many academics have recently been inspired to focus on the digital transformation for medical data thanks to information technology. Electronic health records contain enormous amounts or heterogeneous medical data in many healthcare organizations (EHRs). EHR data must be gathered, integrated, cleansed, stored, examined, and interpreted in a timely manner while maintaining accuracy. Making judgments in healthcare systems necessitates the careful analysis for significant volumes of real-time data gathered through sensors and wearable technology. The outcomes of these choices have a big impact on patient health. For instance, the healthcare systems in a smart city may keep an eye on their patients from a distance and take action if they participate in behaviors that could be harmful to their health. Wearable Internet - Of - things (IoT) devices are being used by researchers to look for signals that could indicate various ailments. However, due to their superior performance, machine learning techniques are growing in popularity.

Descriptive, diagnostic, predictive, & prescriptive analytics are the four basic categories in the field of healthcare analytics. The goal of descriptive analytics is to determine a patient's current condition and to produce data in the form of reports, charts, or histograms. This kind of analysis could be carried out using a variety of analytical instruments. Diagnostic analysis investigates on occurrence of events or the factors that cause them, and it relies on clustering algorithms or decision trees to discover reasons for the recurrence of some serious diseases of individual individuals. By creating a proper prediction model, predictive analytics uses various machine learning techniques to forecast unknown events. By recommending practical steps that result in appropriate patient treatments, prescriptive analytics aims to make the best decisions possible. The demand of big data framework with appropriate analysis tools is essential given the growing volume of healthcare data. This goal has been the subject of significant research, with publications focused on infrastructure, data management, database searching, data mining, & data security. In order to analyse large amounts of healthcare data, this study offers a novel method to e-health that enhances IoT, cloud computing, & fog computing utilizing cognitive data mining or machine learning algorithms. Fog computing enables us to do processing of data over distributed fog nodes, which removes the need for centralized health data processing, which is a barrier to large-scale data analysis.

## II. SOURCES OF HEALTH CARE DATA

Qualitative and quantitative data are collected in datasets related to the health care industry. Quantitative information is utilized for comparison and is of a quantitative nature. Weight, age, temperature, and other discrete variables are some examples. Non-numerical qualitative data are typically used to represent health-related issues. Examples include gender or whether a person smokes or doesn't smoke, etc. Scientific and clinical data are two types of data sources in the medical industry. Data from clinical surveys or epidemiologically based information are examples of clinical data. Data pertaining to bench sciences is referred to as scientific data. Primary and secondary data are recorded and gathered in the medical field. The term "primary data" describes the person or group who collected and analyzed the data. Research questions may be used using the acquired data. Secondary research is dependent on primary data, which is the information that is already available and used for other purposes. Questions based on research are addressed using these data.

## III. CHALLENGES FOR IOT IN HEALTH CARE

Reputable IoT companies' main goal is to offer effective and straightforward IoT / data handling facility deployments. It facilitates the creation of data analytics applications, visualisation frameworks, and IoT health applications by designers. Some of the critical capabilities that IoT organizations must be enabled are:

1) *Simple connectivity:* An excellent IoT company must be capable of offering simple device connection, enabling device administration features.
2) *Easy device management:* It makes various assets and resources more readily available, which improves throughput and lowers maintenance costs.
3) *Information ingestion:* A key component of the Internet of Things is intelligent data transformation and storage. Data is ingested from many data sources, and then using data analytics, pertinent information is extracted.
4) *Informative analytics:* For the best decision-making and efficient operations, proper data analysis is crucial. It is employed in real-time analysis and condition monitoring so that appropriate responses can be made. Additionally, an easy interface makes it easier to grasp and more useful.
5) *Reduced risk:* Respond to alerts and stop any activity that has been gathered within the organizations from a unit console.

## IV. LITERATURE REVIEW

Zeng 2021 et al. based on algorithms related to machine learning. Additionally, this study performed training and fitting of health data using random forest or other related methods. Research demonstrates that the method suggested in the publication can advance the classification of health data. The algorithm can offer technical assistance for the better classification of medical data. [1]

Yang 2022 et albased in medical big data was suggested, and experimental research using questionnaire surveys and other approaches is conducted. Combining all data in the experiment's figure, it can be shown that amongst diabetic patients, men are more likely to have diabetes than women are—12.4% for men and 8.4% for women. According to the information in the figure, the prevalence of retinopathy as a whole is 47.5%, and the rate of retinopathy induced by various sources ranges from 5% to 7%.The duration of the diabetes, urine albumin index, glycosylated haemoglobin index, as well as fasting blood glucose level are just a few of the variables influencing this same prevalence of retinopathy through diabetic patients. Various factors also contribute to the an increase in the incidence of retinopathy through diabetics.

The findings indicate that a variety of factors influence the likelihood that diabetes people may develop retinopathy, thus patients should exercise regularly, maintain a healthy diet, and work to avoid developing retinopathy. [2]

Wei 2022 et al. has grown in popularity as a research subject in the artificial intelligence (AI) community in recent years as a result of its major contribution to a number of application fields. With increasing the number and channel as well as the scale of parameters, it acts as decision support inside the application sector through utilising supercomputing power in the big data era to expose the high-level abstract concepts in the original dataset. A heterogeneous data analysis system for medical education is designed and put into use in this work using DL technology. The suggested system uses decision-level fusion to build a data model, analyses heterogeneous data on medical education, proposes and implements voting and weighting methods, and adopts DL technology for modelling.

To implement the improved DL algorithms for such medical college education dataset analysis approach, the decision value was statistically determined. [3]

Babar 2021 et al. discovered how artificial intelligence and machine learning work. The idea of strengthening the accuracy and usefulness of Ai technologies is also gaining popularity as big information and machine learning take off. In the realm of traffic applications, machine learning techniques improve guard safety in risky traffic situations. The current designs face a number of difficulties, with data privacy being the biggest problem for vulnerable motorists (VRUs). The primary cause of pedestrian traffic control failure is improper user privacy handling. The user information are vulnerable to several security and privacy flaws and are therefore at danger.

If an intruder is able to break into the system, exposed data may be maliciously manipulated, manufactured, and misrepresented for illicit purposes. In this paper, a machine learning-based architecture is suggested for effectively analysing and processing massive data in a secure setting.[4]

A customer medical data exchange plan for privacy-preserving computer vision is put forth by Wang 2022 et al. Our approach combines blockchain technology with a trusted computing environment to make sure that users' control and ownership of their data are not compromised when it is shared. It is suggested to employ a blockchain-based noninteractive secret sharing system so that only users and the TEE may decode shared data. At the same time, humans create a method for user auditing of the sharing process. The security analysis demonstrates that the plan guarantees user data privacy and security throughout storage and sharing. [5]

## V. HEALTH CARE WITH BIG DATA CHALLENGES

The challenges can be categorized into two types:

### *A. Issues Related To Policy And Fiscal Factors*

Medical professionals can only be compensated in the money-for-service era if they interact face-to-face with their patients. It serves as a barrier to the development of new technologies that enable patient participation without their actual physical presence. Additionally, there is more opportunity to use current, cutting-edge technologies where needless face-to-face encounters may be avoided as we move further away form direct interaction-based models, if there are greater financial risks associated. Face-to-face patient encounters in these situations can be highly expensive, but using advanced technologies has a positive impact on people's health results.

### *B. Issues Related To Technology*

The status of health data is one of the biggest technical challenges to completing this task. Medical-related data records created by EHR systems are heavily divided into silos based on organisations. The usage of standardised code sets or message structures is used in an effort to handle this transmission of single data records between silos. However, the problem of data fragmentation remains unresolved. Recently, medical professionals have begun to see that the future of the medical industry lay in data aggregation rather than just the sharing of copies of patient records. Only when information can be acquired from many sources, normalised, and resolved with specific patient identities will the data be considered pertinent and meaningful. Two primary advantages of aggregated data.

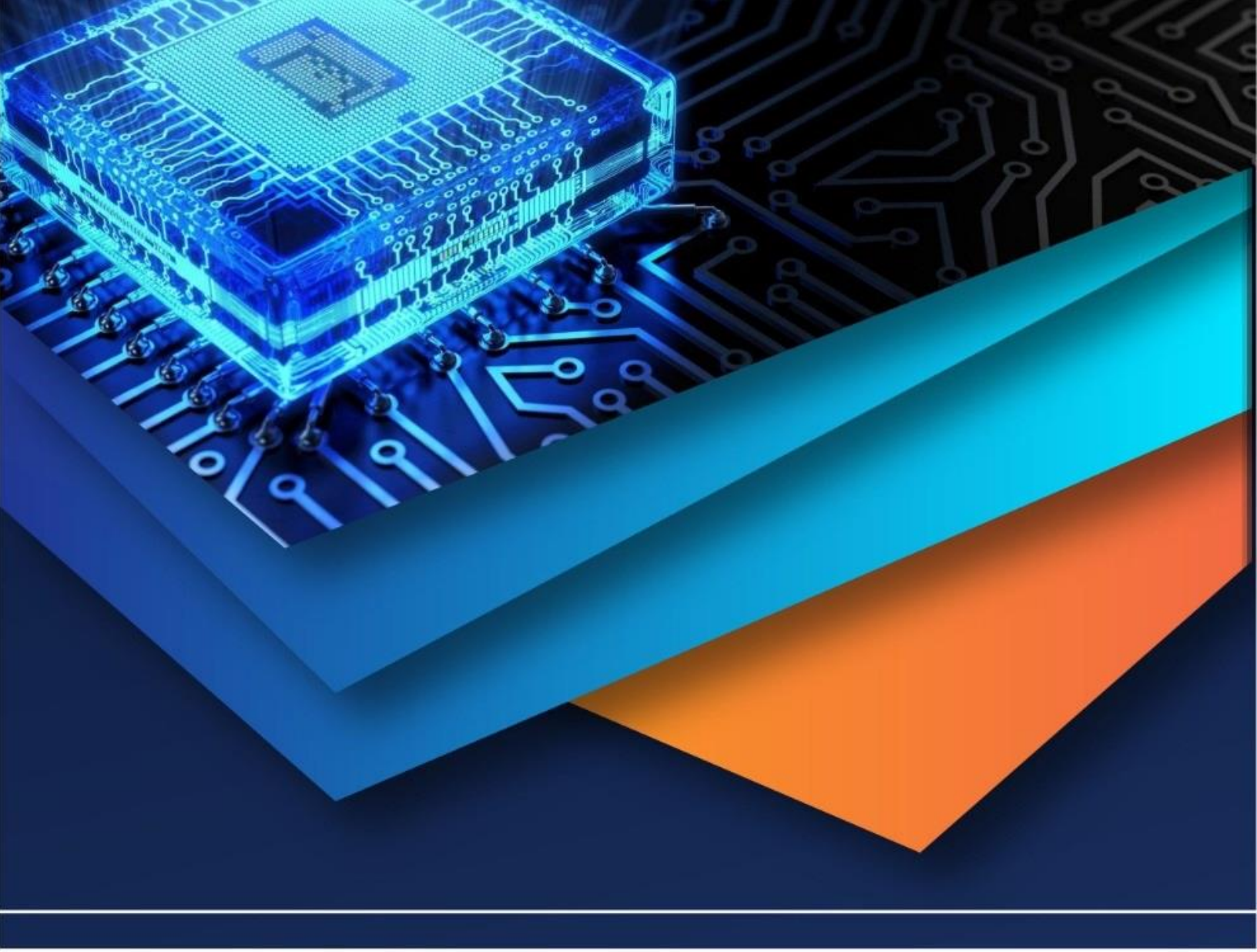
1) The compatibility problem is solved. It is no longer necessary for organisations to create data bridges and transform data between proprietary software. They only have to link sources of data to a shared API module. The foundation for efficient artificial intelligence technologies is this data aggregation.
2) It offers sufficient flexibility, enabling real-time machine learning and artificial intelligence to function well.

## VI. CONCLUSION

There must be a repository where an data is stored in order with an analysis to process it generally. a few instances of repositories and related elements will be provided. Spreadsheets should be mentioned due to their widespread use. They is seen as as the simplest method of data storage. Enabling business analytics to build straightforward logic on rows and columns of data-structured data to provide their own problem analysis. The spreadsheets are very simple to share, and end users are in charge of the logic. But there are also certain drawbacks starting to materialize. Despite their simple sharing and regulated logic, their versioning is a challenge that can be solved. As a result, in terms of sharing, locating the appropriate (the most accurate) version might occasionally be difficult. Furthermore, when a spreadsheets is lost, the data or the logic are also lost because there is no implied backup.

## REFERENCES

[1] Y. Zeng and F. Cheng, "Medical and Health Data Classification Method Based on Machine Learning," J. Healthc. Eng., vol. 2021, 2021, doi: 10.1155/2021/2722854.
[2] L. Yang, Q. Qi, F. Zheng, Y. Wei, and Q. Wu, "Investigation of Influencing Factors on the Prevalence of Retinopathy in Diabetic Patients Based on Medical Big Data," Comput. Math. Methods Med., vol. 2022, 2022, doi: 10.1155/2022/2890535.
[3] L. Wei, Z. Yu, and Z. Qinge, "Medical College Education Data Analysis Method Based on Improved Deep Learning Algorithm," Mob. Inf. Syst., vol. 2022, 2022, doi: 10.1155/2022/3227316.
[4] M. Babar, M. U. Tariq, A. S. Almasoud, and M. D. Alshehri, "Privacy-Aware Data Forensics of VRUs Using Machine Learning and Big Data Analytics," Secur. Commun. Networks, vol. 2021, 2021, doi: 10.1155/2021/3320436.
[5] L. Wang et al., "A User-Centered Medical Data Sharing Scheme for Privacy-Preserving Machine Learning," Secur. Commun. Networks, vol. 2022, 2022, doi: 10.1155/2022/3670107.
[6] Y. Yang, "Correlation Analysis between Sports and Antiaging Based on Medical Big Data," J. Sensors, vol. 2022, 2022, doi: 10.1155/2022/3810676.
[7] J. Li, W. Cui, A. Zeng, Y. Xie, and S. Yang, "Clinical Analysis of Medical IoT and Acute Cerebral Infarction Based on Image Recognition," Mob. Inf. Syst., vol. 2022, 2022, doi: 10.1155/2022/1050264.
[8] Y. Li and Y. Chen, "Research on Chorus Emotion Recognition and Intelligent Medical Application Based on Health Big Data," J. Healthc. Eng., vol. 2022, 2022, doi: 10.1155/2022/1363690.
[9] Y. Huang, S. Nazir, X. Ma, S. Kong, and Y. Liu, "Acquiring Data Traffic for Sustainable IoT and Smart Devices Using Machine Learning Algorithm," Secur. Commun. Networks, vol. 2021, 2021, doi: 10.1155/2021/1852466.
[10] V. D. P. Jasti et al., "Computational Technique Based on Machine Learning and Image Processing for Medical Image Analysis of Breast Cancer Diagnosis," Secur. Commun. Networks, vol. 2022, 2022, doi: 10.1155/2022/1918379.
[11] A. Sriram et al., "A Smart Solution for Cancer Patient Monitoring Based on Internet of Medical Things Using Machine Learning Approach," Evidence-Based Complement. Altern. Med., vol. 2022, pp. 1–6, 2022, doi: 10.1155/2022/2056807.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY